Collaborative scientific platforms for accessing, processing and validation of biodiversity observation and biosensor data.

Francisco Hernandez, Joram Declerck, Klaas Deneudt, Cis Hernandez

Lifewatch - EGI engage May 2015

Outline

- Lifewatch

- EGI - ENGAGE

  - Collaborative scientific platform

  - Biosensors

  - Challenges

  - Status

  - Virtual labs and research environments

- Progress

Lifewatch infrastructure

- European research infrastructure, ESFRI roadmap

- Study of biodiversity and ecosystems

- Access to realtime data

- Biosensor networks

- BIG data, inline data processing

- More info : http://www.lifewatch.eu

- Example VRE : http://marine.lifewatch.eu/

- Belgian contribution: http://www.lifewatch.be

Access

Retrieve and access data resources holding marine biodiversity and ecosystem data. A range of data systems offering data on species names, traits, distribution and genes.

Analyze

Online tools that facilitate data analysis of marine biodiversity and ecosystem data. Analysis is performed on data from known data resources and/or data uploaded by the user.

Develop

Build your own marine virtual lab making use of a range of available web services that access and process data. Service catalogues and 'how to' manuals help you to develop your own system.

Biosensors

- Biological sensors that enable observations at space and time scales relevant to organisms behavior, physiology and life history

    - Optics, Acoustics, Genetics, GPS

- Discrete => continuous

- Delaye

http://www.lifewatch.be/en/sensors
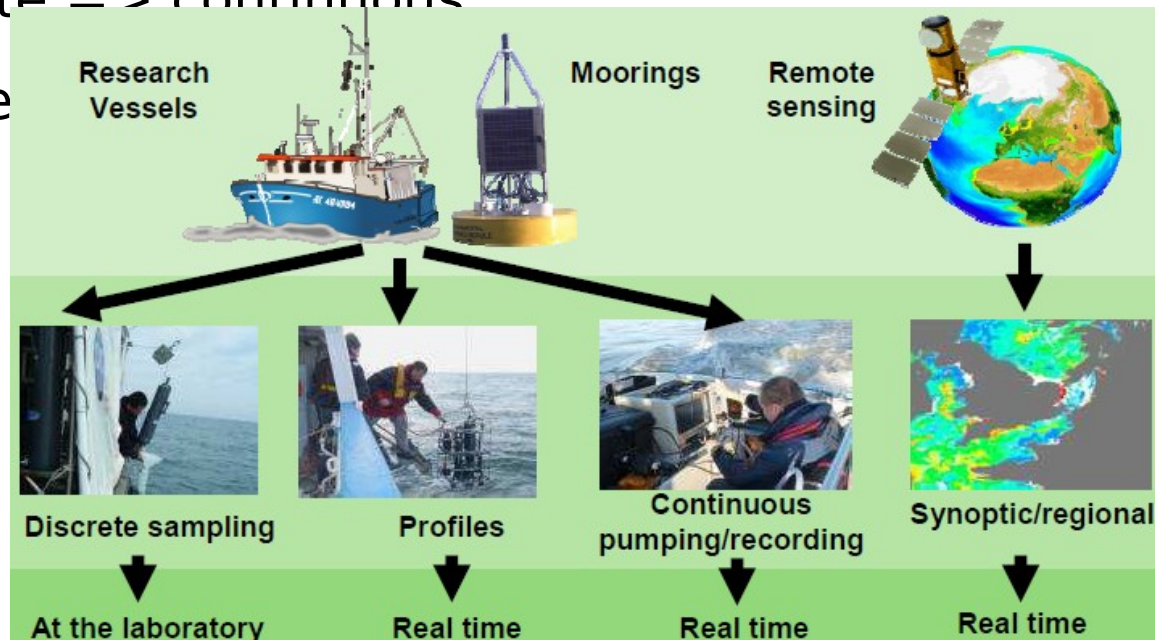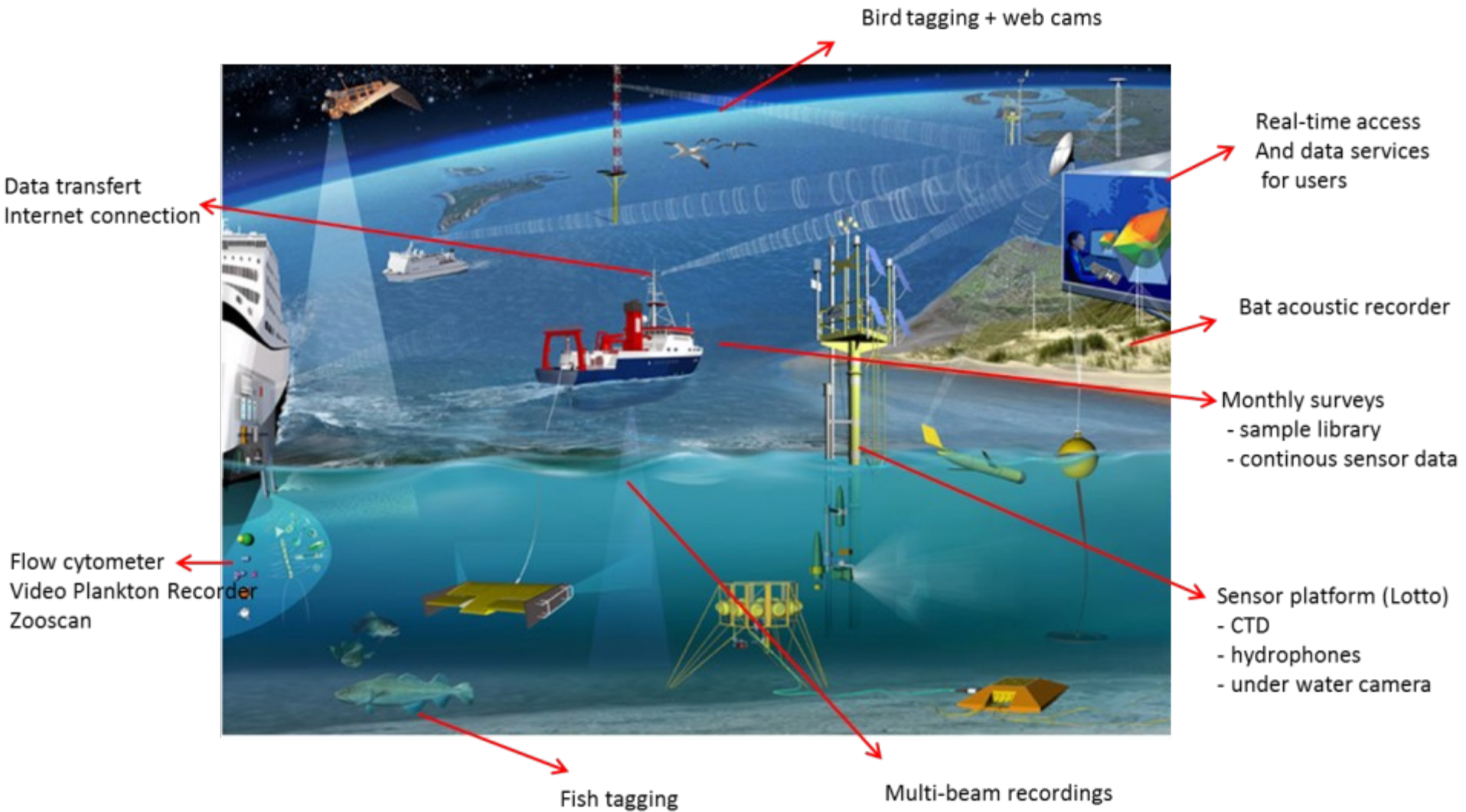
Zooscan

- High resolution flatbed scanner for water samples

- Digital storage and processing of zooplankton samples -> taxonomic composition and abundance

- Image acquisition: 2400-4800 dpi

- Data generation: +/- 4 GB/sample; 432 Gb/year

- Analysis: pattern recognition software (Plankton Identifier [Tanagra], ZooImage [R], …)

Video Plankton Recorder (VPR)

- Real-time underwater digital camera system + strobe

- Rapid quantification of plankton taxonomic composition and abundance

- Image acquisition: 30 frames/second of 7.2 ml image volume

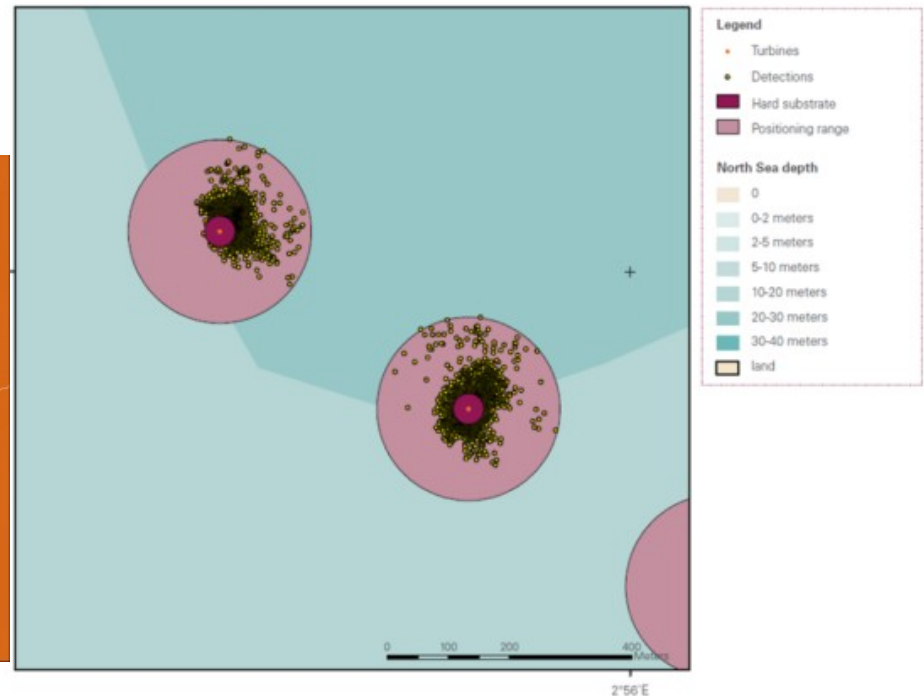- Data generation: +/- 1
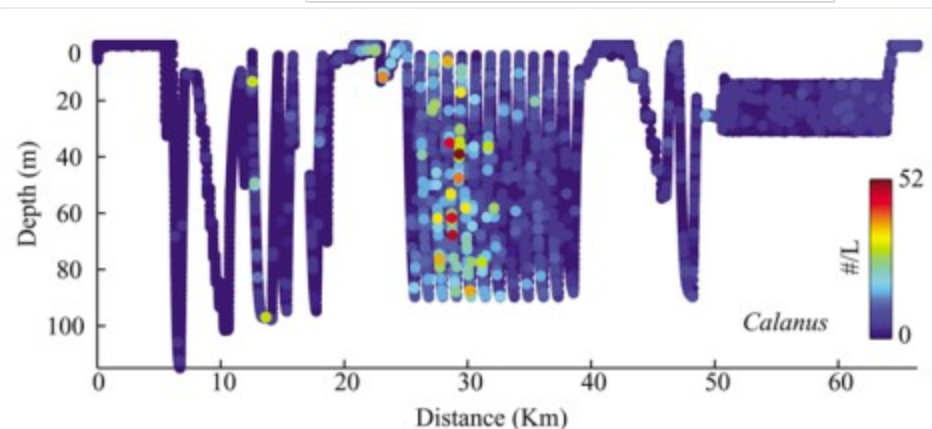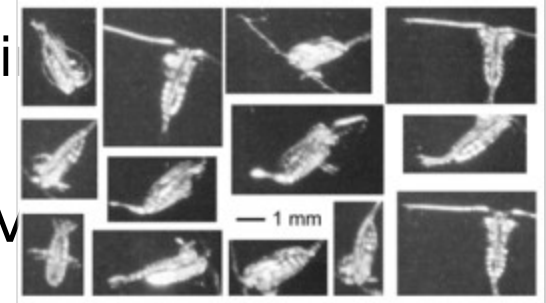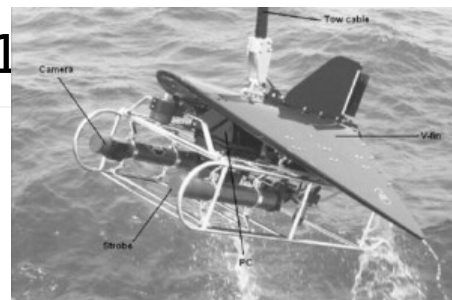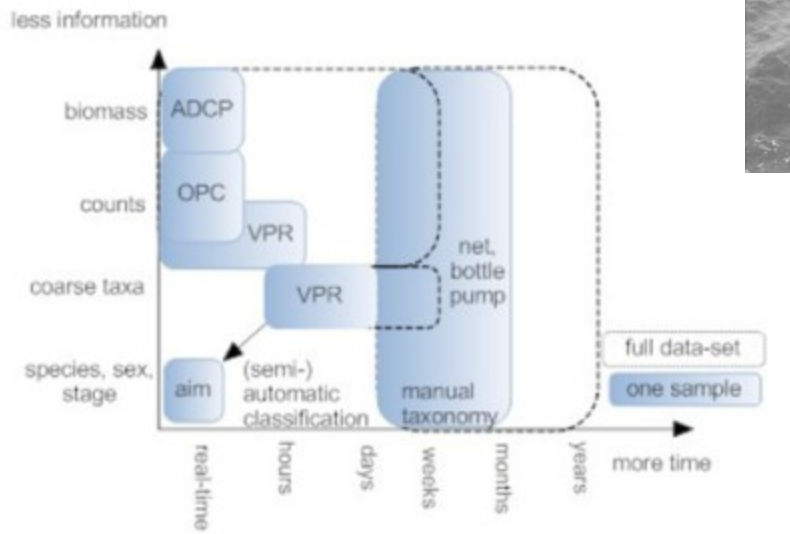
Flow cytometer

- Laser based instrument for particle detection and characterisation in fluids

- Counting and characterisation of phytoplankton particles -> taxonomic composition and abundance

- Image acquisition: particle scattering + fluorescence

- Data generation: +/- 200 MB/sample; 1Tb/year

- ng sof u
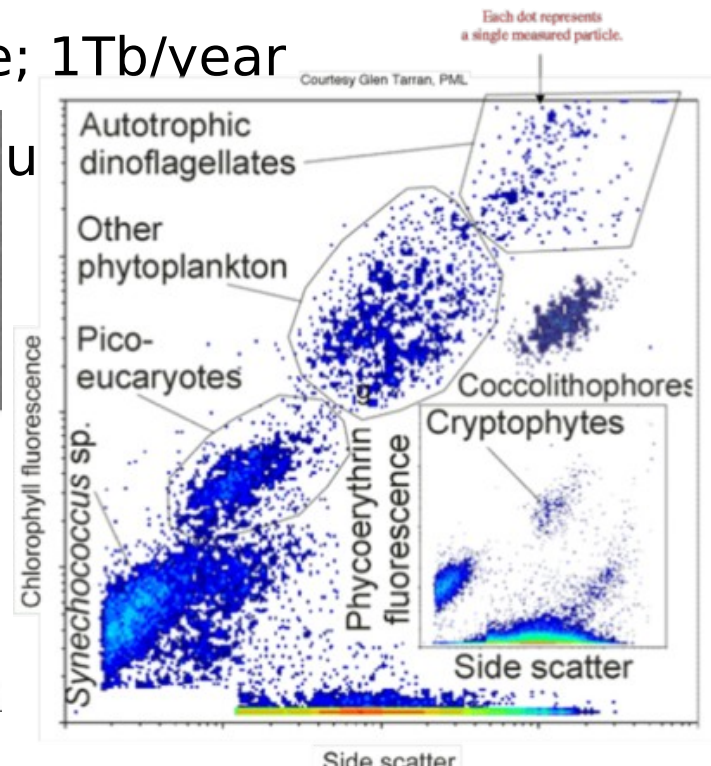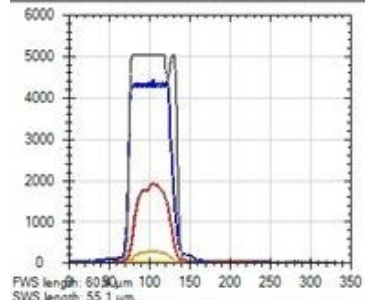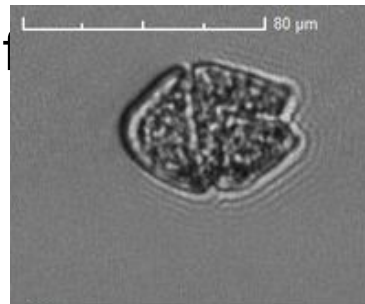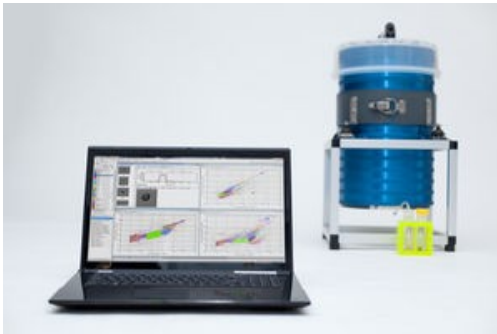
Acoustic fish telemetry

- Acoustic fish tag tracking

- Studying distribution, migration and habitat use

- Data generation: 25 MB/month

- Analysis: GIS mapping & visualization (CartoDB), behaviour analysis (Matlab, Python)

circle represents the area in which position calculation can be performed. The purple circle represents the hard substrate and the yellow dots show the exact positions of the fish.

Multibeam echosounder

- Acoustic high resolution depth sounding sonar

- Bathymetry and sediment typology

- Data generation: sediment 10Gb; water column 100Gb/day

- Analysis: data cleaning and validation, chart creation, deriving sediment typology (CARIS, Fledermaus)
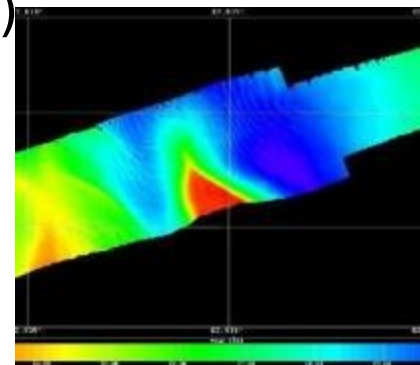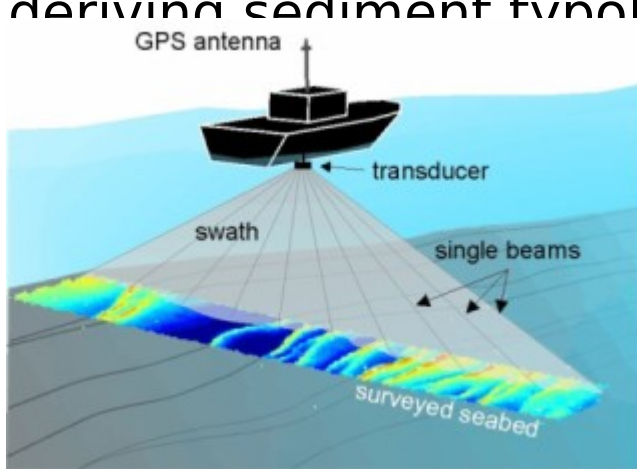
Sediment profiler imaging

- Digital Sediment Profiling Camera

- Vertical cross section of the sediment/water interface

- Image acquisition: 24.1 Mpixel images of 320 cm2 of sediment

- Data generation: 1Gb/image; 130 Gb/year

- Analysis: Image ~~~~~~~~~ti~~ ~f ~~~~~~l~tio~ ~ benthic commu~~~~~~
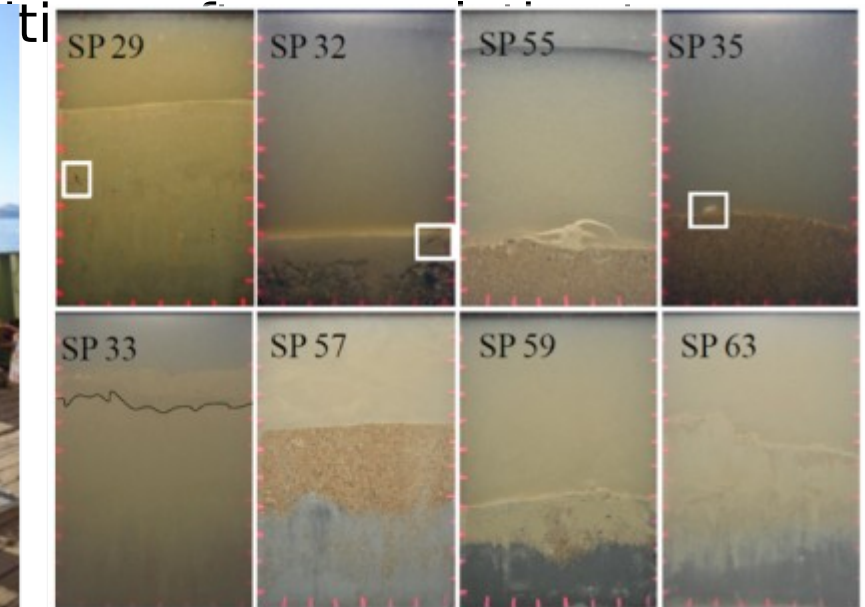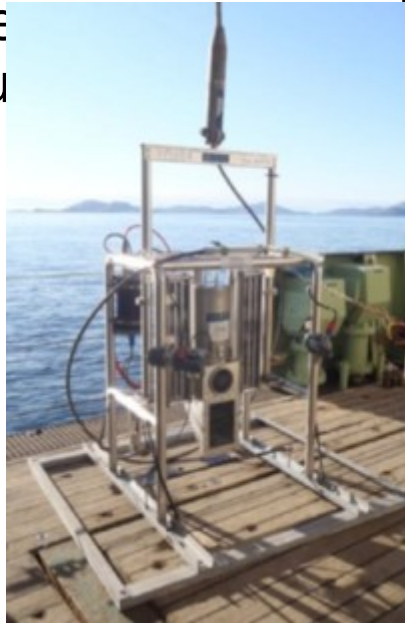
Acoustic bat recorder

- Ultrasound detection and recording

- Sound acquisition: 500 kHz with 16 Bit amplitude resolution

- Data generation: 1 MB/per second of sound recording; 0.5 Gb/night

- Analysis: Call detection and recogition software
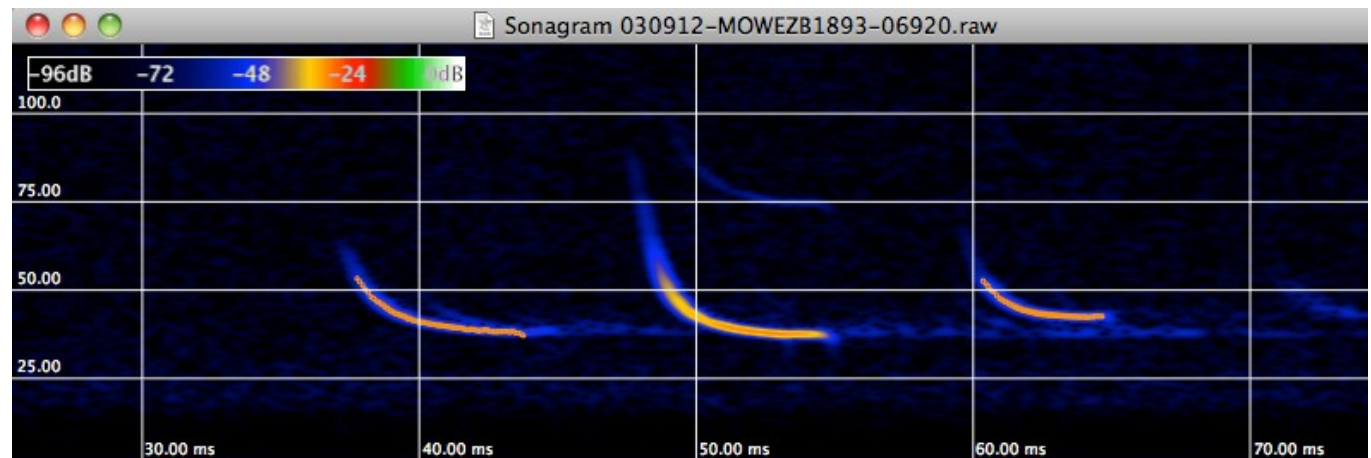
Bird tracking with GPS

- Tracking of large birds with GPS tags developed by UvA-BiTS

- Studying migration and habitat use

- GPS position every 10min, 70 birds, soon > 100

- Data generation: 3G/year (Flemish LifeWatch), multiple GB/year (UvA-BiTS)

- Analysis: GIS mapping & is
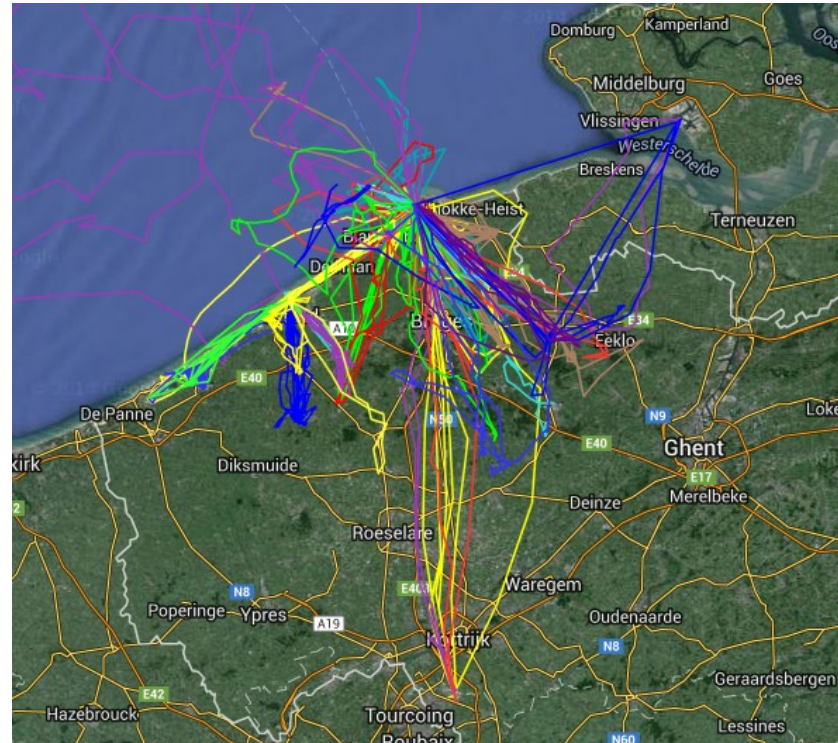
Challenges

- Biosensor
  - buy and install equipment: sensors, winches, antennas
  - technical calibration, raw signal processing, vendor softw.
- IT
  - real-time data transfer, storage, processing
- Biology
  - sensor output -> biodiversity parameters
 - aggregating, combine data, models & predictions, visualization, sharing, validation,…

Doctoral studies on sensor use and application

- 4 PhD's at Ghent University

- Started October 2014

- Work on :

- Standard operation procedures

- Algorithms for translation of sensor output to biodiversity parameters on abundance and distribution of taxa

- Recommendations on optimization and upgrade of the infrastructure

- Applications of the infrastructure in biodiversity and ecosystem studies

Collaborative platform for sensor data processing and analysis
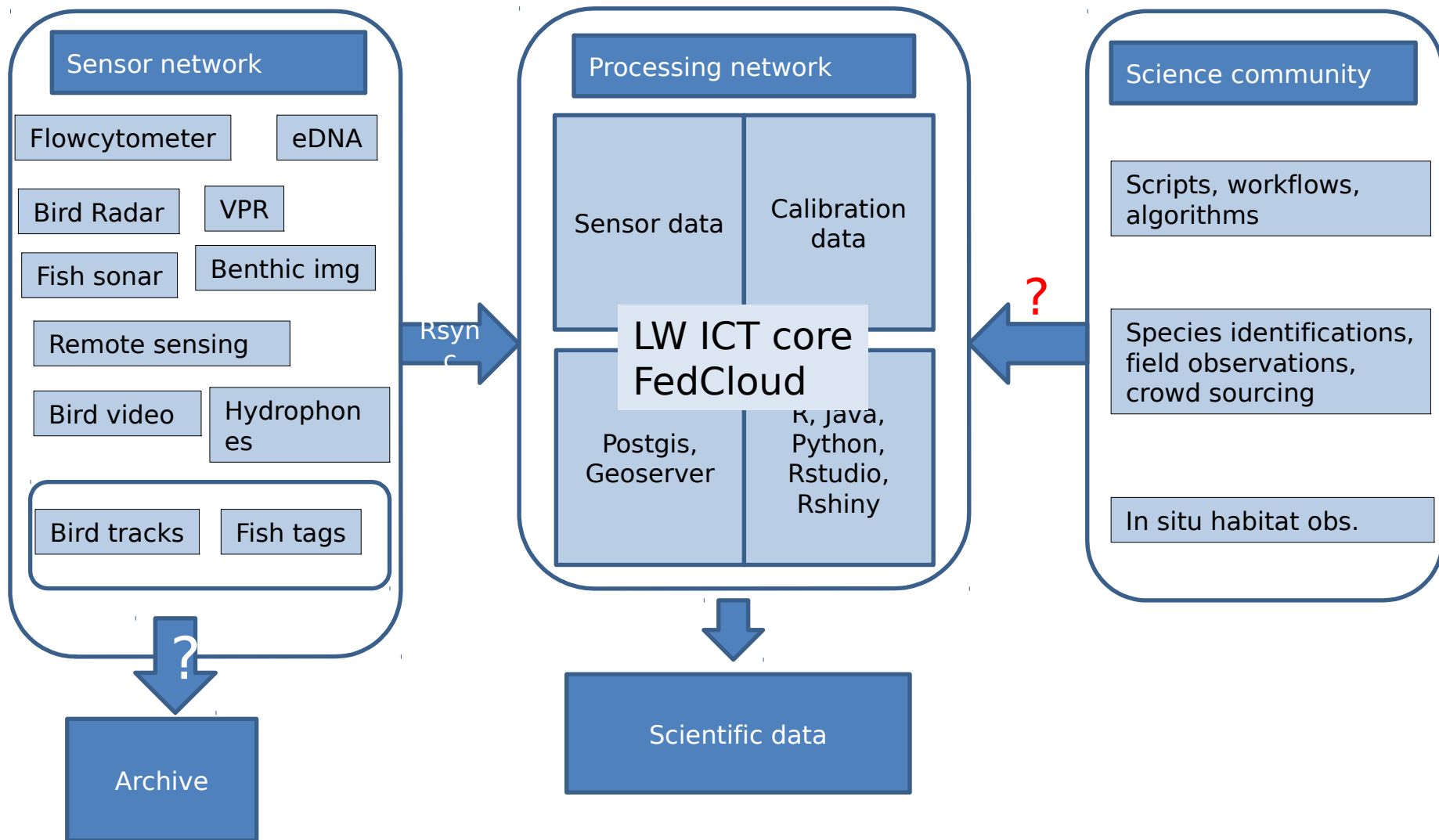
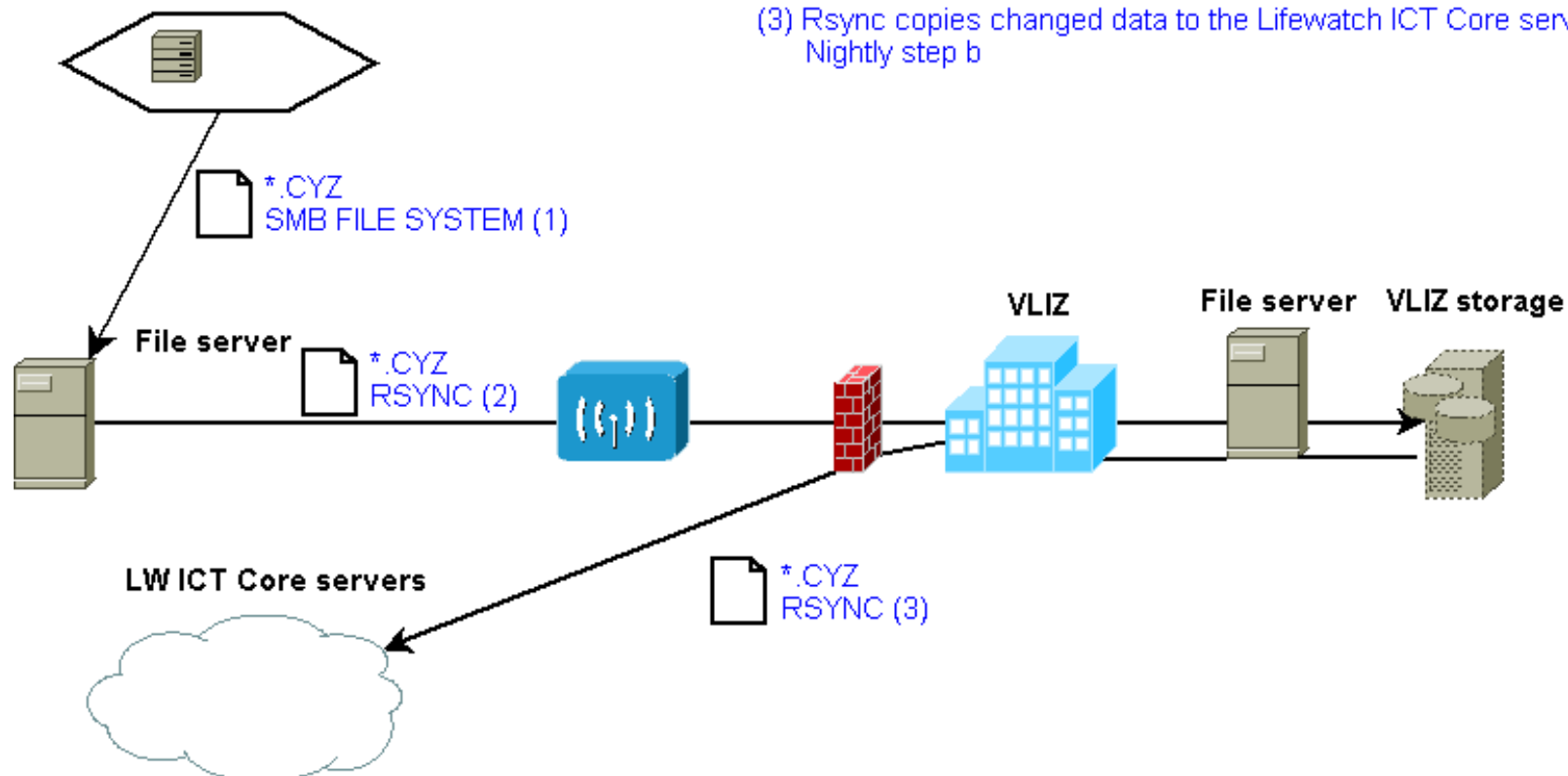| Infrastructure	Status
Bird GPS tracking network + web cams	3 base stations installed and operational 70 birds with GPS
Flow Cytometer	Installed on RV Simon Stevin and operational
Multi beam	Installed on RV Simon Stevin and operational
Acoustic bat detectors	Test installation installed. Offshore installation in preparation
ZooScan	Installed in MSO lab and operational
Video Plankton Recorder	To be installed July 2014
Sediment Profile Imaging	Installed on RV Simon Stevin and operational
Acoustic fish telemetry	40 fishes tagged, 51 receivers installed, data for >2 years

Flowcytometer + embedded PC

(1) Flowcytometer + embedded PC write to shared SMB network filesystem
    Realtime

(2) Rsync copies changed data to VLIZ servers
    Nightly step a

(3) Rsync copies changed data to the Lifewatch ICT Core servers
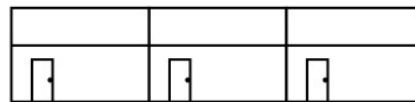   Nightly step b

*.CYZ
SMB FILE SYSTEM (1)

File server *.CYZ
RSYNC (2)

VLIZ **File server** **VLIZ storage**

LW ICT Core servers

*.CYZ
RSYNC (3)

RV Simon Stevin

Video Plankton Recorder

Marine Station Ostend

SSACQ (Fileserver)

RSYNC

ZooScanner

mongoDB network

VLIZ - Main Location

Archive (FileServer)

Script

Collections

MongoDB Primary

vpr

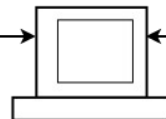ZooSc

MongoDB Replication

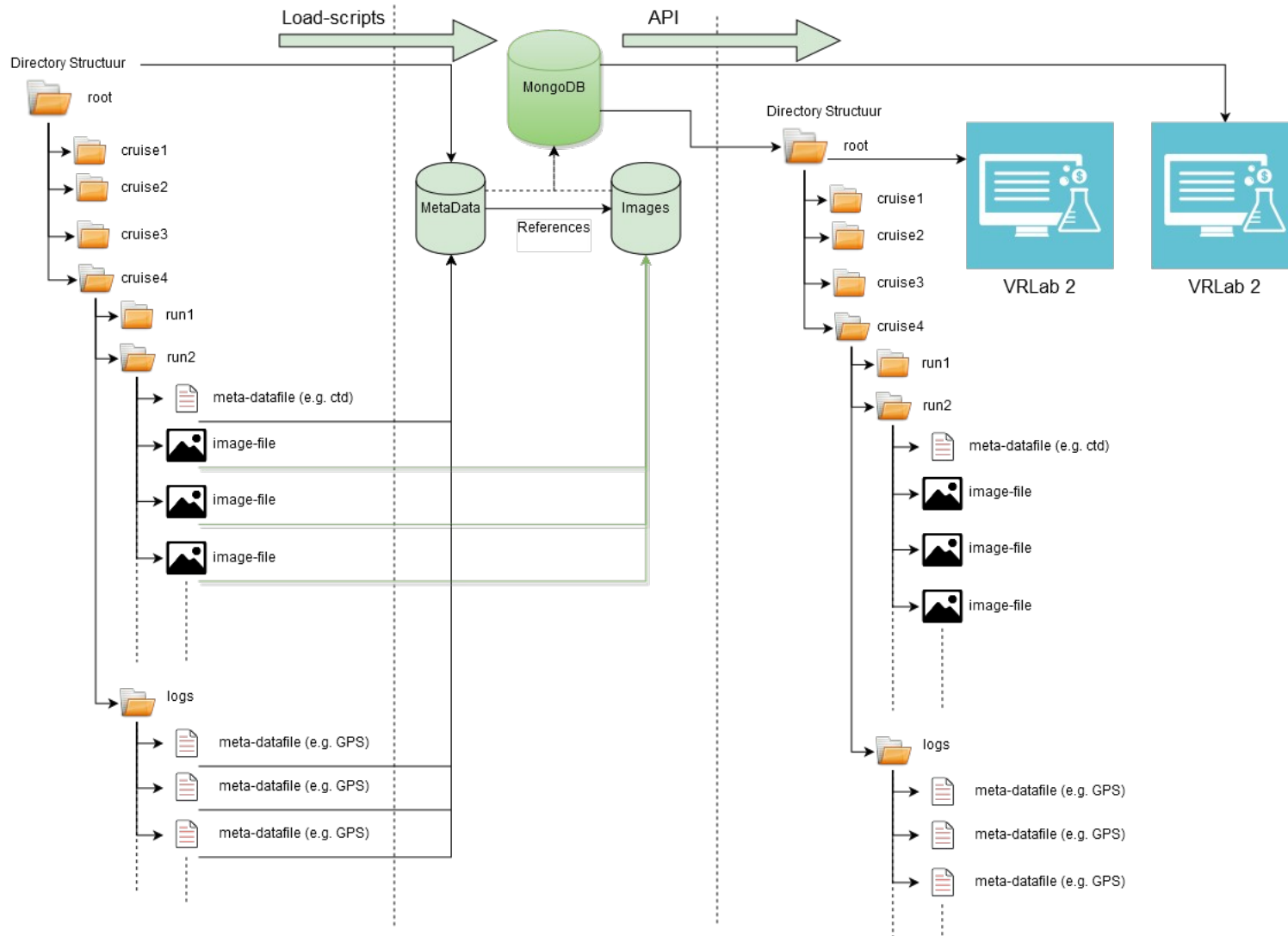Santander - IFCA

MongoDB Secondary
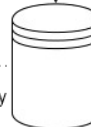
Computing Resources

Data Usage

Virtual Lab

File management

Progress

- Ongoing testing with ICT Core facilities: IFCA, EGI ENGAGE
- Create virtual servers
  - on fedcloud : problem with OCCI
  - on IFCA / LW servers : on going / tickets
  - R, R studio server, Rshiny : at VLIZ
  - PostGIS, geoserver : at VLIZ

- Set up data transfer
  - from ship & stations to IFCA : daily Rsync operational
  - from VLIZ geoserver to IFCA
  - manage storage capacity

- Set up user interfaces for scientists
  - VLIZ users : started
  - external users
  - manage resources: user data, algorithms, scripts, models, …

Lifewatch virtual labs to ICT core

- Meeting in Amsterdam

- Collect and present existing virtual labs
  - Lifewatch Marine Virtual Research Environment (VRE) : done
  - VRE for terrestrial & other thems: started

- Upscale virtual lab components using LW ICT core
  - Environmental layers, Species occurence data
  - BIOVEL portal , Biodiversity catalogue
  - Taxonomic, Biostatical & Modelling webservices

http://marine.lifewatch.eu/

More info

- Lifewatch website [http](http://www.lifewatch.eu)://www.lifewatch.eu
- Marine VRE : http://marine.lifewatch.eu/
- Belgian contribution: http://www.lifewatch.be
- Flanders Marine Institute : http://www.vliz.be